

# Proximal Operator Splitting for Multi-Constraint Dictionary Learning

Zhiyong Liu

Industry Center, Shenzhen Polytechnic, Shenzhen 518055, China  
[zhiyongliu03@163.com](mailto:zhiyongliu03@163.com)

## Abstract

Although the dictionary learning (DL) problem has been extensively studied for about 15 years since the work of Olshausen, the DL problem with multi-constraints on the dictionary atoms has not yet been paid attentions. This paper first explore the DL problem using the newly emergence methods-the proximal splitting methods, such as the iterative shrinkage-thresholding algorithm (ISTA), the fast ISTA (FISTA) and the augmented Lagrange multiplier method (ALMM). Then propose a calculation method, called proximal operator splitting, to split the proximal operator with multi-constraints into several sub-proximal operator. Using this method, the existing proximal splitting methods can be easily extended to deal with the DL problem with multi-constraints. Experiments show that ALMM is a more efficient method than ISTA and FISTA. At last, compare the learned dictionaries of ALMM with the state-of-the-art methods, K-SVD and Majorization. The experimental results show that ALMM outperforms K-SVD and Majorization for correctly chosen constraints.

## 1. Introduction

In recent years, sparse representations have received a lot of attention from the signal processing community. This is due in part to the fact that an important variety of signals such as audio and natural images can be well approximated by a linear combination of a few elements (atoms) of some (often) redundant basis, usually called dictionaries.

Suppose we have a set of signal examples. One of the key problems in sparse representation is to find a dictionary that give rise to the given signal examples via sparse combinations, *i.e.*,

$$\arg \min_{D, X} \varphi(D, X) = \frac{1}{2} \|Y - DX\|_F^2 + \lambda_x R_x(X) \quad (1)$$

where  $R_x(X)$  is a sparsity-inducing regularizer on  $X$ . Typically, for finding a joint sparse approximation of the observations  $Y$ , one can consider a mixed-norm over the rows of  $X$ :

$$R_x(X) = \left( \sum_{i=1}^N \|X_{i, \cdot}\|_q^p \right)^{1/p}$$

where typically  $p=1$  and  $q \in \{2, \infty\}$  [1, 2]. Here, we denote the  $i$ th row and the  $j$ th column of a matrix  $U$  as  $U_{i, \cdot}$  and  $U_{\cdot, j}$ . When the sparsity profile is known to different for each signal to be approximated, we can consider  $p=1$  and  $q=1$  [3] which allows  $R_x(X)$  to be decoupled. Since the work of [4-7], the priors of the dictionary  $D$  are considered. As noted in [5, 8], two typical constraints are the unit Frobenius-norm and the unit column-norm constraints, both of

which lead to nonconvex solution sets. Actually, more constraints, such as sparsity [9] and total variation (TV), can be imposed on  $D$ . Simultaneously imposing multi-constraints on  $X$  and  $D$  leads to the following dictionary learning (DL) problem

$$\arg \min_{D, X} \varphi(D, X) = \frac{1}{2} \|Y - DX\|_F^2 + R_X(X) + R_D(D) \quad (2)$$

where  $R_X(X) = \sum_{t=1}^{T_x} \lambda_X^t R_X^t(X)$  and  $R_D(D) = \sum_{t=1}^{T_D} \lambda_D^t R_D^t(D)$ . Recently, algorithms dealing with multi-constraints or regularized have been proposed in image processing [10-12]. However, few works have addressed this situation for the DL problem. Recently, Yaghoobi *et al.* [9] have proposed a majorization algorithm for DL. Owing to a clever majorization, their algorithm can easily deal with a single constraint over dictionary atoms. However, their algorithm cannot deal with the multi-constraints over dictionary atoms. Recently, the proximal splitting methods have been paid many attentions [10, 12-14]. In this work, we explore the DL problem with the proximal splitting methods and we show how to deal with the multi-constraints problem for these methods.

The paper is organized as follows. Section 1 introduces the proximal splitting methods, such as the iterative shrinkage-thresholding algorithm (ISTA), the fast ISTA (FISTA) and the augmented Lagrange multiplier method (ALMM) and proposes a calculation method, called proximal operator splitting, to split the proximal operator with multi-constraints into several sub-proximal operator. Numerical experiments are presented in Section 2. Section 3 is conclusions.

## 2. Proximal Splitting Methods for Dictionary Learning

Since there are two variables  $X$  and  $D$  in the DL problem, we naturally choose the alternate optimization scheme [5, 8, 9, 15-17] to solve Eq. (2). The alternate optimization scheme involves a signal sparse coding step

$$X^{(k+1)} = \arg \min_X \frac{1}{2} \|Y - D^{(k)} X\|_F^2 + R_X(X) \quad (3)$$

and a dictionary optimization step

$$D^{(k+1)} = \arg \min_D \frac{1}{2} \|Y - DX^{(k+1)}\|_F^2 + R_D(D) \quad (4)$$

Note that the two optimization steps have similar forms, and can be generalized as

$$\arg \min_U f(U) + g(U) \quad (5)$$

where  $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$  is a smooth and convex function, and  $g: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$  is bounded from below but not necessarily smooth nor convex. Since  $g(\cdot)$  might not be smooth nor convex, it is still difficult to solve Eq. (5) directly. Most of the existing methods try to solve it by transforming it to the proximal operator

$$\text{prox}_g C = \arg \min_U \frac{1}{2} \|U - C\|_F^2 + g(U) \quad (6)$$

since, for several forms of  $g(\cdot)$ , the solution of the proximal operator can be computed in closed-form or can be numerically computed efficiently.

Clearly, for the signal sparse coding problem of Eq. (3), we have

$$f(X) = \frac{1}{2} \|Y - D^{(k)} X\|_F^2, \quad g(X) = R_X(X) = \sum_{t=1}^{T_x} \lambda_X^t R_X^t(X) \quad (7)$$

$$D^{(k+1)} = \arg \min_D \frac{1}{2} \|Y - DX^{(k+1)}\|_F^2 + R_D(D)$$

for the dictionary optimization problem of Eq. (4), we have

$$f(D) = \frac{1}{2} \|Y - DX^{(k+1)}\|_F^2, \quad g(D) = R_D(D) = \sum_{t=1}^{T_D} \lambda_D^t R_D^t(D) \quad (8)$$

The form of  $f(X)$  and  $g(D)$  requires the proximal operator  $\text{prox}_g C$  should be extended to a more general form

$$\text{prox}_{g_1, \dots, g_T} C = \arg \min_U \frac{1}{2} \|U - C\|_F^2 + \sum_{t=1}^T g_t(U) \quad (9)$$

In this section, we first discuss how to transform Eq. (5) to Eq. (6) for the iterative shrinkage-thresholding algorithm and the augmented Lagrange multiplier method; we then consider how to split the proximal operator  $\text{prox}_{g_1, \dots, g_T} C$  into several sub-proximal operators.

### 2.1. Iterative Shrinkage-Thresholding Algorithm

Since the only difference between Eq. (5) and Eq. (6) lies in the first function, the iterative shrinkage-thresholding algorithm (ISTA) [14, 18, 19] directly transforms  $f(U)$

to  $\frac{1}{2} \|U - C\|_F^2$  using its second order Taylor approximation:

$$\begin{aligned} & \arg \min_U f(U) + g(U) \\ &= \arg \min_U f(U') + \text{tr} \left( (U - U')^T \nabla f(U') \right) + \frac{1}{2} \text{tr} \left( (U - U')^T \nabla^2 f(U') (U - U') \right) + g(U) \\ &\approx \arg \min_U \text{tr} \left( (U - U')^T \nabla f(U') \right) + \frac{\alpha}{2} \|U - U'\|_F^2 + g(U) \\ &= \arg \min_U \frac{\alpha}{2} \|U - C\|_F^2 + g(U) \end{aligned}$$

where  $C = U' - \frac{1}{\alpha} \nabla f(U')$ ,  $\nabla^2 f(U')$  is approximated by a diagonal matrix  $\alpha I$ . Here,  $\alpha$  is a key factor and can be approximated by the Barzilai-Borwein equation [20]. Then, the optimization problem of Eq. (5) can be iteratively solved by the following proximal operator:

$$U^{(k+1)} = \text{prox}_g \left( U^{(k)} - \frac{1}{\alpha} \nabla f(U^{(k)}) \right) \quad \text{Q (10)}$$

Nevertheless, ISTA has a sublinear convergence rate of  $O(1/k)$ . Nesterov *et al.* [21] and Beck *et al.* [22] propose a fast ISTA (FISTA), which achieves an accelerated non-asymptotic convergence rate of  $O(1/k^2)$ . FISTA can be summarized as the following iterative procedure

$$U^{(k)} = U^{(k)} + \frac{t^{(k-1)} - 1}{t^{(k)}} (U^{(k)} - U^{(k-1)}) \quad (11)$$

$$U^{(k+1)} = \text{prox}_g \left( U^{(k)} - \frac{1}{\alpha} \nabla f(U^{(k)}) \right) \quad (12)$$

$$t^{(k+1)} = \frac{1}{2} \left( 1 + \sqrt{4(t^{(k)})^2 + 1} \right) \quad (13)$$

Substituting Eq. (7) and Eq. (8) respectively into Eq. (10), we have the ISTA for DL problem. Substituting Eq. (7) and Eq. (8) respectively into Eq. (11)- Eq. (13), we have the FISTA for DL problem.

### 2.2. Augmented Lagrange Multiplier Method

Different with ISTA, the augmented Lagrange multiplier method (ALMM) tries to split the objective problem to two sub-problems: one only includes  $f(U)$ , and the other one only includes  $g(U)$  and has the form of the proximal operator  $\text{prox}_g C$ .

To achieve this, ALMM first splits  $f(U)$  and  $g(U)$  by variable splitting scheme, *i.e.*, Eq. (5) is reformulated as

$$\arg \min_{U,V} f(U) + g(V) \quad s.t. \quad U = V \quad (14)$$

Then, ALMM transforms the constrained problem to the following unconstrained one:

$$\arg \min_{U,V} f(U) + g(V) - tr(\Lambda^T (U - V)) + \frac{\mu}{2} \|U - V\|_F^2 \quad (15)$$

where  $\Lambda$  is a dense matrix of Lagrange multipliers related to the equality constraints and  $\mu$  is a parameter that balances the quadratic penalization. Note that Eq. (15) can be rewritten as

$$\arg \min_U f(U) + g(V) + \frac{\mu}{2} \|U - V - \Lambda/\mu\|_F^2 \quad (16)$$

Let  $W = \frac{1}{\mu} \Lambda$ . Then, ALMM solves Eq. (16) using the alternate optimization scheme:

$$U^{(k+1)} = \arg \min_U f(U) + \frac{\mu^{(k)}}{2} \|U - V^{(k)} - W^{(k)}\|_F^2 \quad (17)$$

$$V^{(k+1)} = \arg \min_V g(V) + \frac{\mu^{(k)}}{2} \|U^{(k+1)} - V - W^{(k)}\|_F^2 \quad (18)$$

where  $W^{(k)}$  and  $\mu^{(k)}$  can be updated by  $W^{(k+1)} = W^{(k)} + V^{(k+1)} - U^{(k+1)}$  and  $\mu^{(k+1)} = \rho \cdot \mu^{(k)}$  ( $\rho > 1$  is a constant).

Clearly, Eq. (17) and Eq. (18) have the same forms with the proximal operator. Substituting Eq. (7) and Eq. (8) respectively into Eq. (17) and Eq. (18), we have the ALMM for DL problem.

### 2.3. Proximal Operator Splitting

The core issue of ISTA and ALMM is to solve the proximal operator. Since  $g(U) = \sum_{t=1}^T g_t(U)$ , we next consider how to solve the general form  $prox_{g_1, \dots, g_T} C$  (see Eq. (9)) of the proximal operator  $prox_g C$  by splitting  $prox_{g_1, \dots, g_T} C$  into several sub-proximal operators.

To achieve this, we first reformulate Eq. (9) as

$$\arg \min_{U, V_1, V_2, \dots, V_T} \frac{1}{2} \|U - C\|_F^2 + \sum_{t=1}^T g_t(V_t) \quad s.t. \quad U = V_1 = \dots = V_T \quad (19)$$

The constraint  $U = V_1 = \dots = V_T \in \mathbb{R}^{m \times n}$  can be written as  $HU = V$ , where  $H = [I, I, \dots, I]^T$  and  $V = [V_1^T, V_2^T, \dots, V_T^T]^T$ . Then, we have

$$\arg \min_{U,V} \frac{1}{2} \|U - C\|_F^2 + \sum_{t=1}^T g_t(V_t) \quad s.t. \quad HU = V \quad (20)$$

Using the augmented Lagrange multiplier method (ALMM), we can reformulate Eq. (20) as a non-constrained optimization problem:

$$\arg \min_{U,V} \frac{1}{2} \|U - C\|_F^2 + \sum_{t=1}^T g_t(V_t) - tr(\Lambda^T (HU - V)) + \frac{\mu}{2} \|HU - V\|_F^2 \quad (21)$$

According to the ALMM procedure Eq. (17) and Eq. (18), we have

$$U^{(k+1)} = \arg \min_U \frac{1}{2} \|U - C\|_F^2 + \frac{\mu^{(k)}}{2} \|HU - V^{(k)} - W^{(k)}\|_F^2 \quad (22)$$

$$V^{(k+1)} = \arg \min_V \sum_{t=1}^T g_t(V_t) + \frac{\mu^{(k)}}{2} \|HU^{(k+1)} - V - W^{(k)}\|_F^2 \quad (23)$$

Clearly, Eq. (17) has a closed-form solution

$$U^{(k+1)} = \frac{1}{\mu^{(k)}T + 1} \left( C + \mu^{(k)} \sum_{t=1}^T (V_t^{(k)} + W_t^{(k)}) \right) \quad (24)$$

where  $W_t^{(k)}$  denotes the  $t$ -th block of matrix  $W^{(k)}$ . Since

$$\begin{aligned} V^{(k+1)} &= \arg \min_V \sum_{t=1}^T g_t(V_t) + \frac{\mu^{(k)}}{2} \|HU^{(k+1)} - V - W^{(k)}\|_F^2 \\ &= \arg \min_V \sum_{t=1}^T g_t(V_t) + \frac{\mu^{(k)}}{2} \sum_{t=1}^T \|U^{(k+1)} - V_t - W_t^{(k)}\|_F^2, \\ &= \sum_{t=1}^T \arg \min_{V_t} \left( g_t(V_t) + \frac{\mu^{(k)}}{2} \|U^{(k+1)} - V_t - W_t^{(k)}\|_F^2 \right) \end{aligned}$$

Eq. (18) can be splitted into  $T$  sub-proximal operators, *i.e.*

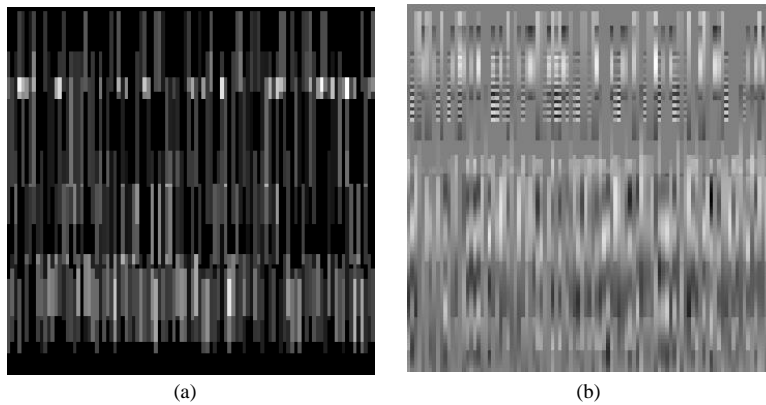
$$V_t^{(k+1)} = \text{prox}_{g_t} \left( U^{(k+1)} - W_t^{(k)} \right) \quad (25)$$

Then, the proximal operator of Eq. (9) can be solved by the union of Eq. (24) and Eq. (25).

### 3. Experiments

As in previously reported works [9, 16], we try the ISTA, FISTA and ALMM algorithm on synthetic signals, to test the ability of these algorithms in recovering the original dictionary that generated signals and to compare the results with other reported algorithms in [9, 16], *i.e.*, Majorization and K-SVD.

Two special types of dictionaries are considered: a dictionary which atoms are piecewise non-zero on a random located segment, called *constant dictionary*, and a dictionary which entries are portion of a cosine function with random frequency, called *cosine dictionary*. Figure 1 shows the signals created by these dictionaries.



**Figure 1. Signals Produced by (a) Constant Dictionary and (b) Cosine Dictionary**

#### 3.1. Comparing Algorithm Efficiency

$$D^{(k+1)} = \arg \min_D \frac{1}{2} \|Y - DX^{(k+1)}\|_F^2 + R_D(D)$$

Since the DL problem of Eq. (3) and Eq. (4) can be solved by an ISTA, FISTA or ALMM method, we expect they provide similar learned dictionaries but they differ in computational efficiency.

We consider the following setting. The dictionary size is set to  $100 \times 10$ . Then, 50 data signals of dimension 100 and  $SNR = 5\text{dB}$  are produced, each created by a linear

combination of 3 different dictionary atoms, with uniformly distributed i.i.d. coefficients in random and independent locations. The regularization of  $X$  is set as  $R_x(X) = 0.05 \sum_{i,j} |X_{i,j}|$ . Several types of constraints, such as unit column-norm, sparsity (SP), non-negative constraint (NN) and total variation (TV), are imposed on the dictionary elements. For all experiments, the dictionary atoms are always constrained to be  $\ell_2$  unit-norm in addition to the other constraints.

For this comparison, we have run these three different algorithms 30 times with different random draws of the dictionary  $D$  and  $X$ . Average runtime (seconds) of ISTA, FISTA, and ALMM for solving the DL problem using different kinds of constraints/regularizers on the atoms from two types of signals are shown in Figure 2. The Regularizers-axis denotes the applied constraints/regularizers in addition to the unit-norm constraint. SP(0.05) denotes sparse regularizer with  $\lambda = 0.05$ ; NN denotes non-negative constraint; TV(0.1) denotes total variation regularizer with  $\lambda = 0.1$ . As shown in Figure 2, with the number of regularizers increasing, ISTA and FISTA cost much more runtime than ALMM. The main reason might be that ALMM using the second-order information while ISTA and FISTA only use the first-order information.

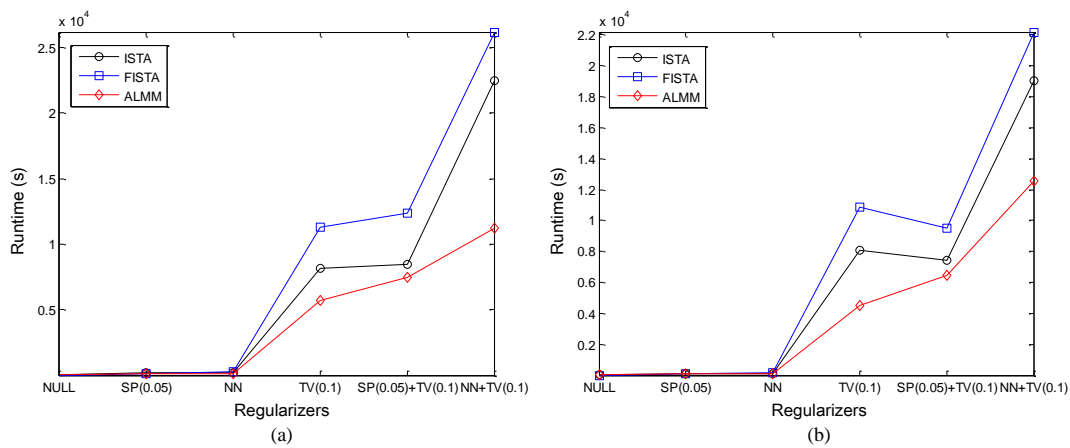


Figure 2. (a) Constant Signals (b) Cosine Signals

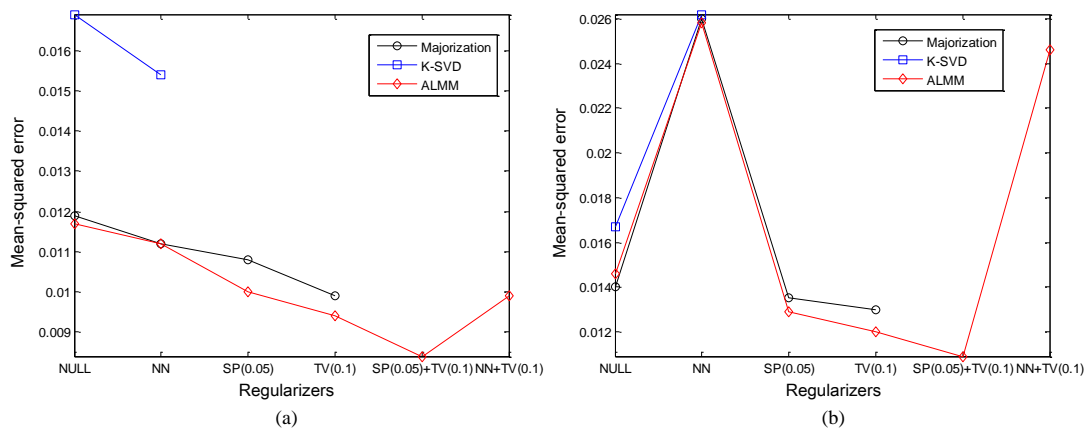
### 3.2. Comparing Learned Dictionaries

In this experiment, we aim at showing the benefit of adapting the constraints on the dictionary to the problem at hand. The experiment shows that considering multi-constraints other than the classical unit-norm constraint clearly leads to better learned dictionary atoms.

The dictionary size is set to  $100 \times 100$ . Then, we produce 400 data signals with dimension 100 and  $SNR = 3\text{dB}$  and each signal is a linear combination of 3 different dictionary atoms, with uniformly distributed i.i.d. coefficients in random and independent locations. Among the 400 signals, we choose 200 signals as training data and the left 200 signals as validation data. For the given value of  $\lambda_x \in \{0.01, 0.05, 0.1, 0.5\}$ , we solve Eq.(2) using the training data to learn the dictionary  $D$ . From the learned dictionary and the same  $\lambda_x$ , we try to approximate the validation signals by solving problem of Eq. (1). The best learned dictionary and its corresponding  $\lambda_x$  are chosen as those that minimize the mean-square error between the validation signals, and their approximations. Finally, for evaluating this best learned dictionary, we solve problem of Eq. (1) using some 5000 noise-free signals and compare the resulting signal approximations

to the true noise-free signals. This procedure has been run 30 times with different draws of the true dictionary and the sparsity patterns.

Figure 3 reports the performance of the K-SVD, Majorization, and ALMM using different constraints. ISTA and FISTA are not considered here since they achieve the same results as ALMM. Except for the unit column-norm constraint imposed on dictionary atoms, K-SVD can only deal with non-negative constraint, and Majorization can only deal with single constraint. Consequently, Figure 3 only reports part of the performances of K-SVD and Majorization. Indeed, we note that when the denoising regularizers, such as TV, or the sparse regularizers, are used, the resulting MSEs of ALMM are lower than the ones of the Majorization method by a ratio of more than 10%. Besides, as expected, using an unsuited regularizer considerably harms the performances (see the non-negative constraint on the smooth dictionary atoms).



**Figure 3. Mean-Squared Error of the Learned Dictionaries of Majorization, K-SVD and ALMM from Two Types of Signals: (a) Constant Signals; (b) Cosine Signals. A Unit-Norm Constraint is Always Considered in Addition to the other Specified Regularizers**

#### 4. Conclusions

We explore the DL problem with multi-constraints through proximal splitting methods, such as ISTA, FISTA and ALMM. We show that all of these methods try to simplify the objective optimization problem into the mature proximal operator problem. Then, we propose the proximal operator splitting method to split the proximal operator with multi-constraints into several sub-proximal operators. Based on this method, the existing proximal splitting methods can be easily used to deal with the DL problem with multi-constraints. Experiments show that ALMM outperforms not only the other proximal splitting methods (ISTA and FISTA) in time efficiency but also the state-of-the-art methods, K-SVD and Majorization for correctly chosen constraints.

In the future, we should extend this work from the following 3 aspects: i) learning the priori information of the dictionary atoms from the trained signals; ii) dealing with the real-world signal processing problem, such as image processing; iii) online learning of the dictionary with multi-constraints.

#### Acknowledgements

This paper is supported by Science Research Project of Shenzhen (No. JCYJ20120613114733545, JCYJ20150630114140642) ,Science Research Project of Shenzhen Polytechnic (No.601422K20007, 601522k35010).

## References

- [1] S.F. Cotter, "Sparse solutions to linear inverse problems with multiple measurement vectors", *IEEE Transactions on Signal Processing*, vol. 53, no. 7, (2005), pp. 2477-2488.
- [2] J.A. Tropp, "Algorithms for simultaneous sparse approximation", Part II: Convex relaxation, vol. 86, no. 3, (2006), pp. 589-602.
- [3] J. Chen and X. Huo, "Sparse representations for multiple measurements vectors in an overcomplete dictionary", In *IEEE International Conference on Acoustics Speech Signal Processing*, (2005), pp. 257-260.
- [4] J.F. Murray and K. Kreutz-Delgado, "An improved FOCUSS-based learning algorithm for solving sparse linear inverse problems", *Signals, Systems and Computers*, (2001), pp.347-351.
- [5] K. Kreutz-Delgado, "Dictionary Learning Algorithms for Sparse Representation", vol. 15, no. 2, (2003), pp. 349-396.
- [6] K. Kreutz-Delgado and B.D. Rao, "FOCUSS-based dictionary learning algorithms", in *Wavelet Applications in Signal and Image Process*, (2000), pp.459-473.
- [7] K. Engan, B.D. Rao and K. Kreutz-Delgado, "Frame design using FOCUSS with method of optimal directions (MOD)", in *Norwegian Signal Process. Symp.*, (1999), pp.214-219.
- [8] K. Engan, S.O. Aase and J. Hakon Husoy, "Method of optimal directions for frame design", in *Proc of the International Conference on Acoustics, Speech and Signal Processing*, (1999), pp. 2443 - 2446.
- [9] M. Yaghoobi, T. Blumensath and M.E. Davies, "Dictionary learning for sparse approximations with the majorization method", *IEEE Transactions on Signal Processing*, vol. 57, no. 6, (2009), pp. 2178-2191.
- [10] M.V. Afonso, J.M. Bioucas-Dias and M.A.T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization", *IEEE Transactions on Image Processing*, (2010), vol. 19, no. 9, pp. 2345-2356.
- [11] C. Chaux, J.P. Pesquet and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems", *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, (2009), pp. 730-762.
- [12] N. Pustelnik, C. Chaux and J.C. Pesquet, "Parallel proximal algorithm for image restoration using hybrid regularization", *IEEE Transactions on Image Processing*, (2011), pp. 2450-2462.
- [13] P.L. Combettes and J.C. Pesquet, "Proximal splitting methods in signal processing", (2011), pp. 185-212.
- [14] P.L. Combettes and V.R. Wajs, "Signal recovery by proximal forward-backward splitting", *Multiscale Modeling and Simulation*, (2006), vol. 4, no. 4, pp. 1168-1200.
- [15] J. Mairal, "Online learning for matrix factorization and sparse coding", *The Journal of Machine Learning Research*, (2010), pp. 19-60.
- [16] M. Aharon, M. Elad and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation", *IEEE Transactions on Signal Processing*, vol. 54, no. 11, (2006), pp. 4311-4322.
- [17] R. Jenatton, "Proximal methods for sparse hierarchical dictionary learning", In *Proceedings of International Conference on Machine Learning*, (2010), pp. 487-494.
- [18] I. Daubechies, M. Defrise and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint", vol. 57, no. 11, (2004), pp. 1413-1457.
- [19] S.J. Wright, R.D. Nowak and M.A.T. Figueiredo, "Sparse reconstruction by separable approximation", *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, (2009), pp. 2479-2493.
- [20] J. Barzilai and J.M. Borwein, "Two-Point Step Size Gradient Methods", *IMA Journal of Numerical Analysis*, vol. 8, no. 1, (1988), pp. 141-148.
- [21] N. Yu, "Gradient methods for minimizing composite objective function", *Mathematical Programming*, (2007), pp. 125-161.
- [22] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems", *SIAM Journal on Imaging Sciences*, (2009), vol. 2, no. 1, pp. 183-202.



## Author



**Zhi-Yong Liu**, he received the Ph.D. degree in mathematics from Sun Yat-Sen (Zhongshan) University, Guangzhou, China, in 2014.

He is currently an Associate Professor of Industry Center, Shenzhen Polytechnic, Shenzhen, China. His current research interests include image processing, wavelet analysis, gait recognition and statistical data analysis.

